PROFESSIONAL

Arson, statistics and the law: can the defendant's proximity to a large number of fires be explained by chance?

E BØLVIKEN

Institute of Mathematics, University of Oslo, PO Box 1053 Blindern, N-0316 Oslo, Norway

and

T EGELAND

Norwegian Computing Center, PO Box 114 Blindern, N-0314 Oslo, Norway

Science & Justice 1995; 35: 97–104
Received 13 January 1994; accepted 23 December 1994

The statistical evidence which played a major part in a case of arson in Norway is presented as a case study. A fireman was known to have been present at the scenes of fire in the hours prior to their onset in no less than 24 out of 37 cases of forest fire. The study, through probabilistic analysis, attempts to throw light on whether this was so strikingly often that he had to be the arsonist, carefully taking into account special features that could explain the peculiar behaviour of the defendant. The conclusion hinged on certain input parameters to the calculation, and the principal aim of the work was to organize, structure and reduce the material to a few quantities that were easier to comprehend than the problem in its original form. The court accepted the relevance of the calculations, and used it against the defendant, but he was still acquitted. A number of issues related to probabilistic interpretation of evidence are discussed.

La preuve statistique qui a joué un rôle majeur dans un cas d'incendie criminel en Norvège est présenté comme un cas d'école. Un pompier dont la présence sur les lieux de feus dans les heures qui précédaient leur allumage était connue dans pas moins de 24 de 37 cas de feus de forêt. L'étude par une analyse de probabiliste, en tenant compte avec soin de caractéristiques spéciales qui pouvaient expliquer l'attitude particulière de l'accusé, tente d'éclaircir si cette fréquence est si choquante qu'il devait être l'auteur des incendies. La conclusion a reposé sur certains paramètres introduits dans le calcul et le but principal du travail a été d'organiser, structurer et réduire le matériel à peu de quantité plus facile à comprendre que le problème dans sa forme originale. La cour a accepté la relevance des calculs et les a utilisé contre l'accusé, mais il a été acquitté malgré tout. Un certain nombre de questions liées à l'interprétation probabilistique de la preuve sont discutées.

Am Beispiel eines Waldbrandes in Norwegen wird eine Wahrscheinlichskeitsberechnung vorgelegt, mit der versucht wurde zu klären, ob das aufällige Verhalten eines bestimmten Feuerwehrmannes diesen als möglichen Brandstifter besonders belastet. Der Tatverdächtige Feuerwehrmann war bei 24 von 37 Waldbränden in der Zeit unmittelbar vor Brandausbruch an den Brandorten gesehen worden. Das Ergebnis der Betrachtung wird von mehreren Parametern beeinflußt. Die bekannten Fakten mußten deshalb sorgfältig abgewogen, strukturiert, geordnet und zusammengefaßt werden. Auf der Basis dieser Datenreduzierung erfolgte eine Wahrscheinlichkeitsberechnung, deren Ergebnis vom Gericht als belastend gewertet wurde jedoch nicht zur Verurteilung führte. In diesem Zusammenhang werden eine Reihe weiterer Gesichtspunkte der wahrscheinlichkeitstheoretischen Beurteilung von Beweismitteln diskutiert.

Se presenta como caso de estudio un caso de incendio provocado en Noruega donde la evidencia estadística jugó un papel fundamental. Se sabía que un bombero había estado en los lugares de los incendios unas horas antes de tener lugar éstos, en no menos de 24 casos de un total de 37 casos de incendios forestales. A través del análisis probabilístico este estudio trató de averiguar si esta presencia era tan frecuente que el individuo tenía que ser el incendiario, valorando cuidadosamente rasgos especiales que pudieran explicar el extraño comportamiento del acusado. La conclusión se basó en ciertos parámetros de partida del cálculo y el principal objetivo del trabajo fué organizar, estructurar y reducir el material a unas pocas cantidades que fueran más fáciles de comprender que el problema en su forma original. El tribunal aceptó la relevancia de los cálculos y los utilizó contra el acusado, aunque fué declarado inocente. Se discuten una serie de posibilidades en relación con la interpretación probabilística de la evidencia.

Key Words: Statistics; Arson; Probability

Introduction

Probabilistic arguments have sometimes been used to assist courts of law in assessing evidence in criminal trials. The delicacy of this task was demonstrated in the much-cited Collins case, see for example [1]. Similarly, the case-study reported here was not clear-cut.

For a number of years there had been an excessively large number of fires in a small rural community in the southern part of Norway. Suspicion of pyromania gradually developed against the second in command of the local fire squad. He stood trial in June 1993, accused of arson in the woods surrounding the community. The charge concentrated on 37 cases of forest fires during a six-week period of drought in the previous year. Technical findings established arson in some cases, but there was no hard evidence directly linking the defendant to the crimes. The material presented by the prosecution was thus only circumstantial. The main evidence against the defendant, from information provided by him during police interrogation, was the fact that he was known to have been at the sites of the fires, shortly before their onset, in no less than 24 of the 37 cases. This information was not disputed by the defence.

It might at first glance seem highly unlikely that proximity in so many instances could be due to chance. The probability under a binomial model would be microscopically small if the defendant was accidentally close in, for example, one of three fires. It would be simple to claim that since a probability of proximity of 1/3 per fire seems high, the defendant cannot possibly have been at the fire sites so regularly by chance. However, such a crude analysis is not good enough in a serious criminal lawsuit when special circumstances might favour the defendant.

Firstly, it was quite a small community. Moreover, the defendant's job of maintaining the roads took him all over the area, stopping and leaving the car for reasons of work. Added to this, he had a medical problem that forced him into the woods much more frequently than the average person. In doing so, could he have been near the subsequent fires without having had anything to do with them?

The first author was called by the prosecution to develop a quantitative argument integrating the special features mentioned above. His report was later scrutinized by a second statistical expert witness, nominated by the defence, who was in general agreement with the work done. The analysis is presented here. In deciding to publicize a solution related to a case of limited public interest, it has been our hope that the mathematical model developed might be put to use in other cases. Moreover, there could be general lessons to learn concerning the modelling, for instance. As always, simplicity should be sought, and as mathematical formulae are likely to be unintelligible to a court, a solution requiring numerical simulation, for example, would be sufficient. The court need not understand technical details in order to act, but it is, of course, all-important that the correctness of the mathematical deductions be unassailably established. What must be kept simple are the issues on which the court is to take a stand, not just its conclusions. It is equally important that a basis is created which allows the court to pass judgement on *assumptions* underlying the analysis and on the values to be assigned to *input parameters*.

In the present case the prosecution and defence held widely differing views on the input parameters. Applied mathematics serves a useful function if it can demonstrate such differences so that the parties can argue and the court decide, after hearing both sides. It follows that a message from a mathematical expert witness should not necessarily be a definite statement. It could also be a summary depending on the assumptions and input values, so that responsibility for issues which are in doubt can be passed on to the court. This requires a bridge between mathematics and law, which in itself suggests that assumptions and input should be kept simple. It may even be worthwhile to sacrifice some accuracy to achieve that goal.

In the present case, the conclusions were phrased in terms of probabilities. A court can be expected to understand, after careful explanation, what a probability of 1:10000 means. The development of the method, and an example of the results presented in court, will be described later. Some authors favour an approach comparing the likelihood of the data given that the defendant is innocent with the likelihood given that the defendant is guilty [2–4]. This must certainly be an appropriate course to follow in many legal applications of probability, although, for reasons set down later, not one followed here.

Regardless of criterion of choice, it might be pertinent in criminal trials to avoid assumptions which do not favour the defendant. The principle of erring in favour of the defendant was followed although this view can be questioned. For example, the present analysis was conducted retrospectively, to some degree obscuring the interpretation of the calculated probabilities.

Furthermore, it is not obvious that mathematical modelling should be carried out with the attitude that the defendant is given the benefit of the doubt at every stage. Perhaps such judgement should be left for the court. These issues and others are discussed later.

The approach

Suppose the defendant has been near the scene of the fire in X out of n (=37) possibilities, and that c (=24) cases of proximity were actually observed. Since nothing is known about the remaining $n-c$ fires, $X \geq c$. If the defendant is innocent, which might mean (although there is no logical implication) that he has appeared at the scenes of fire purely by chance, probabilistic calculus is then available to quantify the plausibility of the observed events. This requires a mathematical model describing in probabilistic terms exactly how the observations (i.e., the $c=24$ proximities) have come about. The model is developed later but this section concentrates on how the results of the analysis should be communicated. The suggested approach is to compute the probability that c or more proximities occur by chance in n trials, or, in mathematical notation,

$$P = \text{Prob}(X \geq c \mid \text{random presence by the defendant}). \tag{1}$$

The formalism implies that the probability is calculated assuming that the presence of the defendant at the various scenes of fire is purely coincidental. P is known in statistics as a p-value and is the most usual way of assessing the outcome of empirical tests of hypotheses. Small values of P signify that the events under consideration are unlikely under the conditions assumed. Thus, one is inclined to disbelieve that chance alone has caused the regular presence of the defendant at the fire scenes if P turns out to be some small number. But, even if the movements of the defendant are judged to have been governed by a mechanism not picking stops randomly with respect to the fires, this can hardly be conclusive proof of his guilt. For one thing, one could imagine enemies of the defendant trying to frame him. Indeed, the weight that should be placed on such a finding is, in our view, a delicate issue, the discussion of which will be deferred until later. For the moment the objective is to set up a probabilistic model based on quantifiable input. Rather than 'exact' values given some mathematical model, upper boundaries on P are derived, permitting substantially simplified model assumptions.

The model

The discussion so far has left a number of vague and imprecise statements which must be put in mathematical form. Consider a specific fire. Imagine the defendant moving around the roads of the community. His speed will be unknown, and sometimes he will not be in the area at all. Occasionally he stops, leaves the car and goes into the woods, possibly close to the site of fire. The more often he does this, the greater is his chance of accidentally closing in on a later fire scene in the relevant time period, so that he *could* have started the fire.

There are three modelling issues to be decided. The first is the number of stops made by the defendant during the critical time period. If T_j is taken as the number of such stops for fire j, T_j can reasonably be regarded as a random variable. Not much will be known about its distribution, being among other things strongly influenced by the length of the time interval over which it is counted. Ideally this should be the time from ignition to the outbreak of the fire, possibly several hours, varying with the way it has been started (candles were discovered at some of the scenes). These details will be avoided later in the paper where mathematical reasoning will establish an upper limit on P which depends solely on the distribution of T_j through its mean.

The second modelling issue is to formalize the meaning of the defendant being 'in the proximity of a fire'. Figure 1 illustrates a definition in which the fire

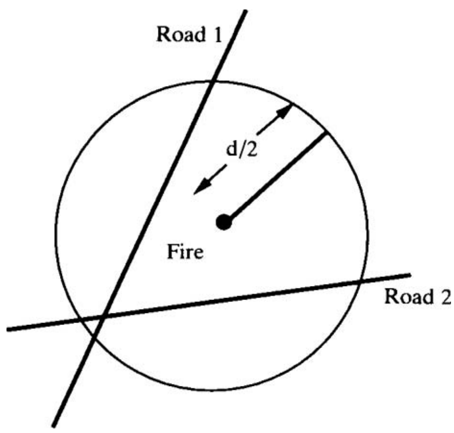

FIGURE 1 Sketch defining neighbourhood to a fire. Two roads are included. If the fires are random with respect to the stopping pattern of the defendant, then Prob(proximity per stop) = l_j/L where l_j is the sum of the chords, and L is the total length of the roads in the community on which the defendant is moving.

in question is allocated to some specific point in the woods, and its neighbourhood is defined by a circle with diameter d. There may be roads passing through, conceivably more than one, as in the figure. Proximity occurs if the defendant happens to stop inside the circle. The value to ascribe to d will have to be selected somewhat arbitrarily. The police suggested a value of 1 km, no doubt retrospectively after having looked at the data, and this was used throughout. It was 2–3 times greater than many of the distances actually observed, which favours the defendant.

It remains to put a value on the probability that the defendant is in the vicinity of the fire each time he stops. This requires two quantities. Let l_j be the local road length adjacent to fire j, i.e., the sum of road lengths inside the defined neighbourhood circle. For example, two roads contribute in Figure 1 and l_j is the sum of the two chords. l_j must be balanced against the size of the area on which the defendant is travelling. L denotes the total road length of the entire community, i.e., the sum of all the roads. It is assumed that the defendant at any time is equally likely to be anywhere on the roads and that his stopping pattern bears no relation to the fire sites. This implies that a stop inside the circle in Figure 1 has probability l_j/L, viz.

$$\text{Prob(proximity per stop)} = l_j/L \tag{2}$$

Later, this equation is converted to a probability $(=p_j)$ of proximity per fire which, in turn, leads to P, which was defined in equation (1).

The values to assign to l_j and L are not obvious. This will be simplified so that only the average, local road length $\bar{l} = (l_1 + \ldots + l_n)/n$ will be required. P is then overestimated. As to the total road length L, about which reasonably accurate information was available from the road authorities, there was, perhaps surprisingly, a considerable difference of opinion between the prosecution and the defence.

Mathematical derivations

This section may be skipped since the subsequent discussion does not rely on it. Most of the mathematics can be found in any elementary textbook on probability and statistics.

The assumptions put forward in the preceding section are enough to derive a useful upper limit on the p-value. First note that the X proximities can be written as a sum of indicator variables, i.e.,

$$X = \sum_{j=1}^{n} I_j \tag{3}$$

where $I_j = 1$ if proximity occurred prior to fire j and 0 otherwise. On the supposition that the defendant is innocent, the independence of $I_1, \ldots, I_n$ can be assumed, so that X is a sum of independent Bernoulli variables with unequal success probabilities, for example, $p_j = \text{Prob}(I_j = 1)$ for fire j. Knowledge is required of the distribution of X at the right tail. Had the values of $p_1, \ldots, p_n$ been known, this would have presented no computational problems. As it is, each p_j is related to the local road length l_j. It did not seem practical to discuss in court the specification of l_j for the 37 different sites of fire.

A simplification was derived by applying an inequality due to Hoeffding [5] (see [6] for an extension), according to which each of $p_1, \ldots, p_n$ can be replaced by their average $\bar{p} = (p_1 + \ldots + p_n)/n$ without decreasing the probability P defined in equation (1). In other words, P is bounded from above by

$$P = \sum_{x=c}^{n} \binom{n}{x} \bar{p}^x (1-\bar{p})^{n-x} \tag{4}$$

provided $\bar{p} \leq (c-1)/n$ ($(=23/37) \approx 0{\cdot}62$ (approx) for the present problem). The condition will be satisfied in all cases of practical interest. Overestimating P by the right hand side of equation (4) is favourable to the defendant.

To use equation (4), $\bar{p}$ must be related to the input quantities from the preceding section. It is enough to derive an upper limit on $\bar{p}$ since the right hand side will increase if $\bar{p}$ is replaced by a larger quantity. First, if $T_j = t_j$ is fixed, and the Bonferroni inequality is combined with equation (2), this yields

$$\text{Prob}(I_j = 1 \mid T_j = t_j) \leq t_j \cdot l_j/L. \tag{5}$$

If the expectation with respect to T_j on both sides is taken, and the rule of double expectation is applied, then

$$p_j = \text{Prob}(I_j = 1) \leq \mu \cdot l_j/L \tag{6}$$

where $\mu = E(T_j)$. Summing over j yields

$$\bar{p} \leq \mu \cdot \bar{l}/L \tag{7}$$

where l is the average local road length as described earlier. P can now be approximated from the above by combining inequalities (7) and (4).

The right hand side of equation (7) may exceed 1 but this causes no problems since the limit is accurate for small values of $\bar{p}$, that is, those of interest. Bonferroni boundaries always behave in such a way. Improvements are available from convexity arguments if an additional assumption is made. Suppose the places

selected by the defendant to stop are independent of each other. Then, using equation (2)

$$\text{Prob}(I_j = 1 \mid T_j) = 1 - (l - l_j/L)^{T_j}. \tag{8}$$

Taking expectation with respect to T_j on both sides, as before, leads to

$$p_j = 1 - E(1 - l_j/L)^{T_j} \le 1 - (1 - l_j/L)^{\mu} \tag{9}$$

after noting that Jensen's inequality can be used to replace the random variable T_j by its mean. Summing over j on both sides and dividing by n yields, after another round of the Jensen inequality,

$$\bar{p} \le 1 - (1 - l_j/L)^{\mu} \tag{10}$$

provided $\mu \ge 1$. This sharpening of equation (7) gives considerable improvements in some of the cases treated in the next section, but the extra assumption lacks foundation and might not favour the defendant.

Numerical results

Table 1 was computed using the Bonferroni based boundary equation (7). Although quite conservative, it avoids the independence condition underlying the alternative boundary equation (10). The validity of the latter and more accurate estimate was considered unproven, and it was not used in court (the methods coincide for $\mu = 1$, but deviate considerably for $\mu > 1$). The court was handed several tables of this kind.

The expected number of stops per fire (μ) and the total road length (L) were varied as shown, the attitude being that it was for the court to decide on appropriate values. L was a matter of dispute. The defence lawyer maintained that the medical problem forcing the defendant to stop in isolated places, far from people, implied a low value for L. The third input parameter, the average local road length $\bar{l}$ was also selected by crude judgement. Clearly $\bar{l}$ is proportional to the diameter d of the circle defining the neighbourhood in Figure 1. It was put at 2 times d, that is 2 kilometres, which is an estimate on the high side, and favoured the defendant.

The general impression conveyed by the table is one of extreme sensitivity with respect to the input parameters. If the average number of stops is taken to be 2, the P-value ranges from $1/10^{13}$ for $L = 50$ km, i.e., 0 for all practical purposes, to $1/(4 \times 10^3)$ for $L = 12{\cdot}5$ km. (Scientific notation is beyond what we expect jurors to understand and is therefore avoided in Table 1.) This sensitivity can not be helped. Users of the results must accept it as a fact of life. It suggests, of course, that the actual numerical p-values should not be taken too literally. The main attention must obviously be directed towards the input parameters, although no one can say exactly which values to supply for them. There is no way other than by common sense, the jury deciding in the end.

TABLE 1 Assessments of the p-value (P) and the average probability proximity per fire ($\bar{p}$), using equation (7)

Average no of stops per fire (μ)	*Total road length* (L) (*km*)	$\bar{p}$	P
1	50	0·04	$<1:10^{13}$
	37·5	0·05	$<1:10^{13}$
	25	0·08	$<1:10^{13}$
	12·5	0·16	$1:10^{10}$
2	50	0·08	$<1:10^{13}$
	37·5	0·11	$1:10^{13}$
	25	0·16	$1:10^{10}$
	12·5	0·32	$1:4\times10^{3}$
3	50	0·12	$1:10^{13}$
	37·5	0·16	$1:10^{10}$
	25	0·24	$1:10^{6}$
	12·5	0·48	$1:3\times10^{1}$
4	50	0·16	$1:10^{10}$
	37·5	0·21	$1:50\times10^{6}$
	25	0·32	$1:4\times10^{3}$
	12·5	0·64	large*
5	50	0·20	$1:10^{8}$
	37·5	0·27	$1:0{\cdot}5\times10^{6}$
	25	0·40	$1:1{\cdot}5\times10^{2}$
	12·5	0·80	large*

* Hoeffding's inequality does not apply.

Discussion

Many authors have recommended probabilistic evidence in courts of law interpreted from the so-called Bayesian perspective [2–4]. There is certainly something to be said for agreeing on a single, general framework for carrying out analyses, and this section points out how the p-value approach relates to the Bayesian one. The discussion is best conducted in terms of odds, which in a criminal trial means the likelihood for innocence divided by the likelihood for guilt. More precisely, if I stands for innocence, G for guilt and E denotes the piece of evidence in question, the odds in favour of innocence becomes

$$\frac{\text{Prob}(E \mid I)}{\text{Prob}(E \mid G)} \times (\text{odds for } I \text{ from other information}), \tag{11}$$

which is a version of Bayes's formula. The formula tells us that a given piece of evidence E should be judged through its likelihood ratio under I and G. The numerator, $\text{Prob}(E \mid I)$ is the probability that E occurs if the defendant is innocent while the denominator $\text{Prob}(E \mid G)$ is the probability that E occurs if the defendant is guilty. Note that the odds for innocence will go up compared to what it was prior to taking E into account if $\text{Prob}(E \mid I) > \text{Prob}(E \mid G)$, i.e., if E is more likely when the defendant is innocent than when he is guilty. If the combined expression in equation (11) equals, say, 0·25, then the odds in favour of guilt are 4:1, or, to put it differently, the likelihood of guilt is 0·8.

All this is elegant and important, but when applied to the present case not very different from what is reported here. Indeed, if E is identified with the event $X \geq c$, it demonstrates that at least c proximities have occurred (see above). Now, $\text{Prob}(X \geq c \mid G) = 1$ (since the defendant appears at every fire scene if he is the arsonist) while $\text{Prob}(X \geq c \mid I)$ is precisely the quantity called P earlier. Hence, the likelihood ratio in equation (11) coincides with the p-value on which the previous analysis has been based. The Bayesian perspective has thus led to an alternative interpretation of P. It could be that the judiciary should be educated to interpret evidence in this way and that all should agree on the Bayesian method to carry out probabilistic analyses in legal contexts, as advocated by Lindley [3] and Robertson and Vignaux [4]. However, in the case described here, this was not the main point. The authors' brief was not to come up with a probability for guilt nor to calculate the second factor of equation (11). It is important that the odds for guilt go up by a factor of 1000 if $P = 1/1000$, and perhaps more could have been made out of this than was. But the central issue, as indeed it must often be, is the uncertainty in the assessment of P itself. This was what the discussion in court had to concentrate on.

Calculations of the sort attempted above invariably run into two problems obscuring the interpretation of the probabilities. First of all, they are computed retrospectively for precisely the reason that the data suggested that they would be small. This implies a selection bias. The probabilities appear smaller than they should (similar quantities are not computed in other cases). Another point is the number of people that could conceivably have behaved similarly to the defendant, i.e., appearing at the wrong place at the wrong time. The more of them there are, the less emphasis should be placed on a coincidence involving a single person. This aspect, although crucial in the Collins case, is of little importance here, and neither do we think the problem of retrospectivity is critical. The overriding concern is the enormous uncertainty arising from ignorance (partial at least) of the input parameters. For one thing, this makes the standard significance levels in statistics too large in the present context. With a sensitivity as shown in Table 1, surely nominal levels of 1% or 0·1% do not make much sense. The attitude must be that to proclaim that the movements of the defendant were governed by chance, the p-values must remain small (less than one in ten thousand or in one hundred thousand, for instance) over the range of input combinations deemed relevant.

The court accepted this sentiment and shared the views of the prosecution on what was the relevant part of Table 1. It concluded, therefore, that the presence of the defendant was unlikely to be due to chance, but there was little corroborating evidence, and he was acquitted. Rejection of the hypothesis of random presence at the fire spots does not logically imply guilt, but it is a clue which the court considered a strong one. Judging and weighing evidence is, of course, the ultimate responsibility of the courts. With the present piece of evidence they may be handicapped since the concept of randomness may not be an easy one to grasp for members of the legal profession and for laymen. The heart of the matter is how strong a presumption of guilt is implied by non-random behaviour. No mathematical formalization seems to be of help here so the appraisal belongs solely to the realm of the judge or jury. Only their assessments count. The dilemma is that they, unlike statisticians, lack deep insight into the meaning of the concept of randomness. No doubt the courts' attitude towards evidence in probabilistic terms will vary, but there must be a danger that some of them will give unreasonable credence to probabilistic analyses. However, this applies equally to many other forms of expert testimony (for example, statements made by psychologists and psychiatrists). Indeed, statisticians do not, in principle, see their role as any different from that of other experts in other trials. The court will have to trust the professional competence of the experts in question, bring in fellow experts to examine their work, struggle to understand their statements and carefully assess the weight to be attached to their evidence. Statistical experts differ only if statistical and mathematical methodology are sought to perform the more ambitious task of actually entering the decision making of the court, as discussed, for

example, by other workers [1, 3, 4] (with differing conclusions).

The concern here has been with what Tribe [1] has called mathematics as a fact finder, although not in a mechanistic sense with the aim of drawing a sharp conclusion. The scope of this work has been the much more limited one of organizing and structuring material by reducing it to a few parameters that are easier to comprehend than the problem in its original form. The merit of the work done hinges on whether it is better for the court to take a stand on the parameters that have been constructed than on the raw data. With the danger of judges and juries over-trusting the analysis noted in addition, it is prudent to ask whether the analysis should have been presented to the court at all. It would not be surprising to learn that opinion was divided among our colleagues. Admittedly, there are weak spots in the analysis, but what is the alternative? The court would find it difficult to interpret the evidence better on its own. The analysis did reduce the problem to a few intelligible key conditions, creating a better basis for argument and decision-making than would otherwise be possible.

A key factor in all expert testimony is that of communication [7]. In itself this suggests analyses that are as simple as possible. In the present case the authors could have gone further in quantification. It would have been possible, for example, to incorporate parameter uncertainty formally into the calculations through so-called Bayesian assumptions on the input parameters. The results might have looked more stable on the surface, but it would not have enhanced the discussion in court. For one thing, it would serve as an extra obstacle to the lawyers and the laymen, but even more importantly, the unclear points about the input, which would then have been buried in mathematical statements, are for the court to decide. It would not have helped if this had to be accomplished through specification of means and variances in prior probability distributions.

Finally, to what extent should mathematical analysis and modelling seek to avoid everything disfavouring the defendant? This issue distinguishes the analysis reported in this paper from a large number of applications in other areas and also from civil lawsuits [8]. It was, in the present case, necessary to simplify input by using upper limits on the p-values. This was conservative and to the advantage of the defendant, but it also served as a trick to obtain estimates of p-values from simplified assumptions. The issue would have been the same with the Bayesian approach. Note that the required sharpness of the boundaries depends on the actual values. In situations with small p-values, the assumptions of the mathematics can be relaxed and generous to defendants whereas more marginal cases will have to be treated more rigorously. The reported Bonferroni boundaries are likely to be quite conservative, but they sufficed and the extra assumption required to sharpen the boundaries seemed hard to justify. That was the opinion of the statistical expert nominated by the defence.

One may wonder how far the dictum of being favourable to the defendant should be carried. Applied mathematical analysis tends to contain some weak spots. A mathematician cannot prove his assumptions, and statistical models describing complex systems usually harbour a degree of uncertainty regarding the correctness of the model. Moreover, it may be maintained that giving the defendant the benefit of the doubt is the privilege and duty of the court and not the expert. According to this view the expert serves best by keeping strictly to what he considers 'correct'. The problem is that it may be hard to decide what is 'correct'. Even if it is up to the court to pass judgment in principle, the problems that statisticians have in mind are within the mathematical modelling, usually beyond reach of the understanding of lawyers and judges. Uncertain parameters can be varied, and the impact inspected by the court, as explained above. But the choice between the two limits produced is something different. The statistical expert simply has to take a stand. Should he do so with the attitude of erring on the side of the defendant, or should he try for what he thinks is closest to the 'truth'? To the authors the answer is not obvious, and the matter will not be pursued.

Acknowledgments

We are grateful for helpful comments from Inspector Trond Dyvik, Professors Inge Helland and Tony Vignaux and, above all, from Dr CGG Aitken. Criticism from referees improved the presentation.

References

1. Tribe L. Trial by mathematics: precision and ritual in the legal process. Harvard Law Review 1971; 84: 1329–1393.
2. Evett I. Interpretation: a personal odyssey. In: The Use of Statistics in Forensic Science. Chichester: Ellis Horwood Ltd, 1991: 9–22.
3. Lindley D. Statistics and the law. The Statistician 1977; 26(3): 203–220.

4. Robertson B and Vignaux G. Expert evidence: law, practice and probability. Oxford Journal of Legal Studies 1992; 12(3): 392–403.
5. Hoeffding W. On the distribution of the number of successes in independent trials. Annals of Mathematical Statistics 1956; 27: 713–721.
6. Boland J and Proschan F. The reliability of k out of n systems. Annals of Probability 1983; 11(3): 760–764.
7. Napley D. Lawyers and statisticians. Journal of the Royal Statistical Society A; 1982: 145(4): 422–438.
8. DeGroot M, Fienberg S and Kadane J. Statistics and the Law. Wiley series in Probability and Mathematical Statistics. Chichester: John Wiley & Sons, 1986.